

Research Article

Restless Watchdog: Selective Quickest Spectrum Sensing in Multichannel Cognitive Radio Systems

Husheng Li

Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996, USA

Correspondence should be addressed to Husheng Li, husheng@eecs.utk.edu

Received 26 January 2009; Revised 29 May 2009; Accepted 8 July 2009

Recommended by K. Subbalakshmi

Selective quickest spectrum sensing, which monitors the spectrum activity in multiple channels, is studied for multichannel cognitive radio systems with nonnegligible channel switching time (blind period). The spectrum sensor needs to detect the emergence of primary users as quickly as possible. Due to hardware limitation, it is assumed that only a subset of frequency channels can be monitored simultaneously. The problem of controlling the monitoring procedure is studied in the frameworks of dynamic programming (DP). System states and cost functions are defined. Cost-to-go functions for DP are derived, simplified, and approximated, based on which control policies are derived. Numerical results are provided to demonstrate the proposed algorithms.

Copyright © 2009 Husheng Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

In recent years, cognitive radio [1, 2] has attracted intensive studies since it helps to solve the underutilization problem of frequency spectrum [3–5]. Significant progress has been made for standardizing (e.g., IEEE 802.22 [6, 7]) and implementing (e.g., XG radio system [8]) the cognitive radio. In a cognitive radio system, secondary users (without license) can use frequency channels that are not being used by primary users (with license). However, when primary users emerge, the secondary users need to quit the frequency channel as quickly as possible. Therefore, spectrum sensing, which monitors the spectrum activity, is a key issue in cognitive radio systems, particularly in wideband systems which may contain multiple frequency channels. Due to hardware limitation (e.g., limited sampling rate), it is difficult to sense all frequency channels simultaneously. A feasible strategy is to monitor only a subset of the frequency channels and hop to another subset (may be the same as the current one) in the next time slot according to a certain control policy. Typically, it requires time to transit between different frequency channels (e.g., the time needed for reconfiguring phase-locked loop (PLL), which typically takes 1 milliseconds, and the time for selecting band-pass

filter, which depends on the design of circuits), during which the secondary user cannot sense any frequency channel (thus we call it *blind period*). Therefore, the spectrum sensor is like a restless watchdog (illustrated in Figure 1), which runs from one door to another to monitor possible intruding thieves (actually, the primary user is the owner of the house. However, from the viewpoint of the secondary user, it is an intruding thief. In contrast to real life, the secondary user handles the thief by quitting the house, instead of dialing 911), and cannot monitor any door when it is running.

The problem of selecting channels to sense in multichannel cognitive radio systems has attracted plenty of researches. In [9, 10], conclusions in bandit problem [11] are applied to study the tradeoff between exploration and exploitation when the channel characteristics are not fully known. Similarly, the framework of bandit problem is also applied in [12], which is focused on finding an indexing policy for different channels. In [13], the framework of Partially Observable Markov Decision Process (POMDP) is applied to choose suitable channels for accessing. The work in [14] has discovered a surprising conclusion that myopic policy is optimal in certain circumstances. Note that the references listed here, although numerous, are far from being exhaustive.

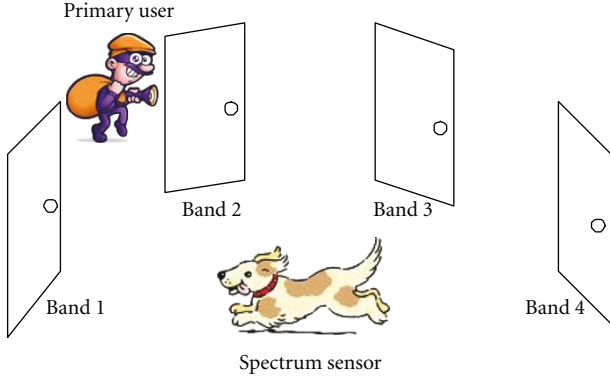


FIGURE 1: Illustrating spectrum sensing over multiple frequency channels.

In this paper, we target at finding an intelligent controlling policy for this restless watchdog. In contrast to the above existing studies, which study the procedure of accessing channels, this paper is focused on the procedure of quitting channels being used by secondary user when primary users emerge. The main concern of this procedure is the delay of detecting primary users (the longer the delay is, the more violation primary users suffer) as well as false alarms. Since the detection of primary users needs to be as quickly as possible, we adopt a framework similar to the quickest detection [15, 16] in this paper. However, the study in this paper is different from the single channel case in [15, 16] since the spectrum sensing in the multichannel cognitive radio system needs to not only detect the primary users as quickly as possible but also select suitable channel(s) to sense. Therefore, we coin the algorithm studied in this paper as *selective quickest spectrum sensing* to distinguish from the proposed quickest spectrum sensing in [15]. Note that the meaning of “sensing” here is more like “monitoring” instead of looking for new opportunities as used in many literatures.

Note that a similar selective quickest spectrum sensing problem has been addressed in [17], which discusses the other side of the story, that is, finding available frequency channels for data communication. Therefore, the incentive of spectrum sensing in [17] is to get reward from locating blank frequency channels while the incentive in this paper is to avoid the penalty of conflicting primary user. In contrast to the restless watchdog in our paper, the spectrum sensor in [17] is more like a food-hunting lion. Different analysis tools are used: theory of partially observable Markov decision process (POMDP) is used in [17] while Dynamic Programming (DP) is applied in this paper. Moreover, this paper considers blind period, which substantially impacts the structure of decision making (e.g., it is difficult to find explicit optimal control policies), while [17] ignores it.

For the controlling policy of the restless watchdog, we try to solve the following two problems based on noisy observations (assuming that only one frequency channel can be monitored at a time).

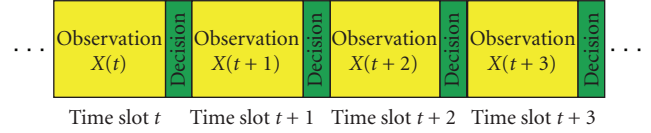


FIGURE 2: Timing structure of the spectrum sensing.

- (i) When to claim the detection of primary users' emergence and stop communicating over the frequency channel being monitored? Note that a good spectrum sensor needs to achieve good tradeoff between detection delay (impacting the communication of primary users) and false alarm (impacting the communication of secondary users themselves).
- (ii) When to switch to another frequency channel that is not being monitored? Which frequency channel should be switched to? Note that the secondary user is blind during the transition period and there exists risk that primary users emerge during this blind transition period.

In this paper, we assume that the emergence of primary users is memoryless. Therefore, the above controlling problem falls in the field of Markov Decision Process (MDP). Naturally, we apply the framework of Dynamic Programming (DP) [18, 19], which provides the optimal solution, to study the above two problems. A brief introduction to DP is provided in Appendix A to make this paper self-contained.

The remainder of this paper is organized as follows. The system model is given in Section 2. Elements of control problems, system state, action space, and cost function, are defined in Section 3. Cost-to-go functions in DP are analyzed for finite and infinite horizon cases in Sections 4 and 5, respectively. The control policy is further simplified using heuristic approximation in Section 6. Numerical results and conclusions are given in Sections 7 and 8, respectively.

Below is some mathematical notation used in this paper.

- (i) For sets A and B , $A/B = \{x \mid x \in A, x \notin B\}$; $|A|$ means the cardinality of set A .
- (ii) $\|\mathbf{x}\|_1$ means the 1-norm of vector \mathbf{x} , that is, $\|\mathbf{x}\|_1 = \sum_k |(\mathbf{x})_k|$; $\|\mathbf{x}\|_0$ means the 0-norm of vector \mathbf{x} , that is, the number of nonzero elements in \mathbf{x} .
- (iii) $(x)^+$ is equal to x if $x \geq 0$ and 0 otherwise.

2. System Model

Suppose that there exist M frequency channels being used by a secondary user. A secondary user needs to sense the frequency spectrum and monitor the activities of primary radios. Once a primary user emerges on a frequency channel, the secondary user needs to vacate from it. We denote by H_0^m (H_1^m) the hypotheses that the m th frequency channel is not being used (is being used) by primary users. The time is slotted and labeled by integers $0, 1, 2, \dots$.

The following assumptions are placed on spectrum sensing.

- (i) At the beginning, all M channels are idle and are being used by the secondary user. (In this paper, idle means that the channel is not being used by primary user.)
- (ii) The activities of primary users on different channels are mutually independent. This is reasonable since different channels are typically assigned to different communication systems or transmission links. It is interesting to study the case of correlated channels; however, it is beyond the scope of this paper.
- (iii) Suppose that the procedure of spectrum sensing is time slotted. At the beginning of each time slot, a new observation on the spectrum activity is received. Then, the decision of action is made at the end of the time slot. We denote the observation at time slot t by $X(t)$ and the observations from time slots t_1 to t_2 by $X_{t_1}^{t_2}$. This procedure is illustrated in Figure 2.
- (iv) Only one frequency channel can be monitored at a time. Switching to another frequency channel needs d_s time slots (the blind period), during which the secondary user cannot sense any channel. We denote by O_m the set of the indices of time slots in which channel m is sensed. By changing the definition of system states, it is easy to extend the result to the case that more than one channels can be monitored simultaneously.
- (v) We assume that the probability distributions of observations, with and without primary users, are perfectly known to the secondary user for all frequency channels. We denote the observation distributions of hypotheses H_0^m and H_1^m by p_{0m} and p_{1m} , respectively. Note that there is no *a priori* information about these distributions in practical systems. However, they can be estimated from the experience of secondary users. For simplicity, we ignore the procedure of learning the information in this paper.
- (vi) Suppose that the emergence time of primary user on a frequency channel satisfies geometrical distribution and the corresponding probability is given by $p_e(t) = \rho(1 - \rho)^{t-1}$, where the subscript e stands for emergence, and we assume that ρ is identical for all frequency channels and is known to the secondary user. In practical systems when the true value of ρ is unknown, we can either estimate it or use an artificial ρ as a parameter to control the agility of spectrum sensing. Note that the assumption of geometrical distribution is identical to the two-state Markov chain assumption [10, 13], where the transition probability from state “idle” to state “busy” is ρ .
- (vii) For simplicity, we do not consider the procedure of finding new available frequency channels. This task can be accomplished by applying the techniques in [17] and can also be easily incorporated into the framework of this paper.
- (viii) We do not consider the case of multiple secondary users, in which competition is unavoidable and makes the control policy much more complicated.
- (ix) For simplicity, we do not consider the period of data transmission and assume that the spectrum sensing is continuous in time. In practical systems, data transmission is carried out orthogonally to the spectrum sensing, either in frequency or in time. When the orthogonality is in frequency, the spectrum sensing can be carried out in a subband of each channel and the data transmission can be done in the remainder of the spectrum (some guard band can be used to prevent frequency leakage) such that spectrum sensing and transmission can exist simultaneously. When the orthogonality is in time, the spectrum sensing and data transmission are carried out in different time slots (like time-division-multiplexing (TDM)). In this case, we can skip the data transmission period when computing the metrics used in spectrum sensing since the data transmission period does not provide information for the spectrum sensing. Therefore, in both cases, we can assume that the spectrum sensing is carried out continuously in time without violating practical system designs.

3. Elements of Control Problem

The selective quickest spectrum sensing is essentially a control problem which generally has three elements: system state, cost function, and action space. The action space is obvious. We will explain the two elements, system state and cost function, for the selective quickest spectrum sensing in this section.

3.1. System State. When $M = 1$ (single frequency channel), the secondary user has only two states, namely, continuing using/sensing the current channel and stop transmitting over this channel. When $M > 1$, the definition of states needs to incorporate the information of frequency channels being used. When at least one channel is being used for transmission, we denote a generic state by S_m^Ω , where Ω denotes the set of channels being used for data communication and $m \in \Omega$ stands for the channel being sensed. When Ω is an empty set, the state, denoted by S^0 , means that all frequency channels have been closed by the secondary user.

Then, the set of all states, denoted by \mathbf{S} , is given by

$$\mathbf{S} = \{S_m^\Omega \mid m \in \Omega, \Omega \subseteq \{1, 2, \dots, M\}\} \cup \{S^0\}. \quad (1)$$

It is easy to verify that the cardinality of \mathbf{S} is given by

$$\begin{aligned} |\mathbf{S}| &= 1 + \sum_{m=0}^{M-1} \binom{M}{m} (M - m) \\ &= 1 + M2^{M-1}, \end{aligned} \quad (2)$$

where 1 stands for the state S_0 , m is the number of closed channels, $\binom{M}{m}$ is the number of possible selections of m

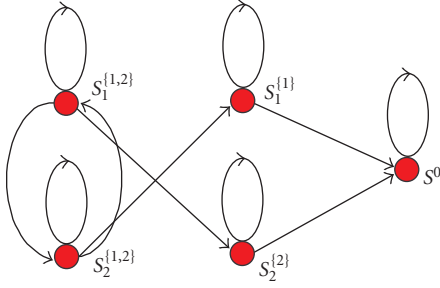


FIGURE 3: Illustration of state transitions when $M = 2$.

closed channels, and $M - m$ is the number of possible selections of channels being sensed.

The spectrum sensing allows transitions from state $S_m^{\Omega_1}$ to state $S_n^{\Omega_2}$ only when $\Omega_2 \subseteq \Omega_1$. When $\Omega_2 = \Omega_1$, the transition means that the secondary user switches from channel m to channel n without stopping transmitting over any channel. When $\Omega_2 \subset \Omega_1$, the transition means that the secondary user stops communicating over channel m and switches to sense channel n .

An illustration of state definitions and transitions is provided in Figure 3 when $M = 2$. Below are two examples of state transitions.

- (i) From $S_1^{\{1,2\}}$ to $S_2^{\{1,2\}}$, the secondary user still continue to use channels 1 and 2 for communication and switches to sense channel 2.
- (ii) From $S_1^{\{1,2\}}$ to $S_2^{\{2\}}$, the secondary user stops using channel 1 and only channel 2 will be used and sensed (the transmission and spectrum sensing may not occur simultaneously as explained in the last section).

3.2. Cost Function. We measure the system performance by false alarms and detection delays. Similar to [20], we consider the following cost function:

$$\begin{aligned}
 J &= \sum_{m=1}^M P(T_m > t_m) + c \sum_{m=1}^M E[(t_m - T_m)^+] \\
 &= \sum_{m=1}^M P(T_m > t_m) + c \sum_{m=1}^M E\left[\sum_{k=1}^{t_m-1} P(T_m \leq k)\right],
 \end{aligned} \tag{3}$$

where T_m is the time slot when primary user emerges in channel m , t_m is the time slot of detecting the primary user and stopping transmitting over channel m , and c is a constant scalar balancing the weights of false alarm and detection delay. In the second equation in (3), we used the equality

$$E[X] = \sum_{k=0}^{\infty} P(X \leq k), \tag{4}$$

where X is a nonnegative random variable. Note that the first summation in (3) means the sum of false alarm probabilities

and the second summation denotes the sum of average run length (ARL) of detection delay in all frequency channels (for channel m , the detection delay ARL is $E[(t_m - T_m)^+]$). Then, in each time slot, the secondary user may experience a false alarm penalty $P(T_m > t_m)$ if claiming detection of primary users on channel m or a miss detection penalty $P(T_m \leq k)$ for channel m if continuing using channel m .

4. Finite Horizon Case

In this section, we consider a finite period of spectrum sensing and use DP to obtain optimal rule of selective quickest spectrum sensing.

4.1. Cost-to-Go Function. As an important tool in DP, cost-to-go function means the expected cost from current time slot to final time slot Γ . The details can be found in [19]. We assume that the spectrum sensing is carried out in a finite interval $[0, 1, \dots, \Gamma]$. At the end of time slot Γ , the secondary user must quit all channels and restart the procedure of finding available channels.

For the finite horizon case, we define the cost-to-go function $J_t(s)$, where t indicates time slot and s indicates state, in a similar manner to [20], which is given by (note that the cost-to-go function is conditioned on observations)

$$\begin{aligned}
 J_t(s | X_0^t) &= \sum_{m=1}^M P(T_m > t_m, t_m \geq t | S_t = s) \\
 &\quad + c \sum_{m=1}^M E\left[\sum_{k=t}^{t_m-1} P(T_m \leq k | S_t = s)\right],
 \end{aligned} \tag{5}$$

where S_t stands for the state at time slot t . Obviously, the cost incurred before time slot t is omitted in $J_t(s)$, and only the cost after $t - 1$ is taken into account.

Following the backward induction of dynamic programming, we begin the discussion from the cost-to-go function at the final time slot Γ . Provided observations X_0^Γ , the cost-to-go function at state S_m^Ω and time slot Γ is given by

$$J_\Gamma(S_m^\Omega | X_0^\Gamma) = \sum_{n \in \Omega} P(T_n > \Gamma | X_0^\Gamma), \tag{6}$$

which is sum of false alarm probabilities at Γ (recall that we need to close all channels at time slot Γ).

For $0 \leq t < \Gamma$, the cost-to-go function for state S^0 is given by $J_t(S^0 | X_0^t) = 0$ since all channels have been closed and there will be no more cost in the future.

For $0 \leq t < \Gamma$ and $|\Omega| \geq 1$, the cost-to-go function for state S_m^Ω is given by

$$\begin{aligned}
 J_t(S_m^\Omega | X_0^t) &= \min\left(C_m^\Omega(m | X_0^t), \min_{n \neq m} C_m^\Omega(n | X_0^t), \min_{n \neq m} \tilde{C}_m^\Omega(n | X_0^t)\right),
 \end{aligned} \tag{7}$$

where the operation of minimization stands for choosing the action incurring the minimum cost. Note that, in (7),

$C_m^\Omega(m | X_0^t)$ is the cost to go for remaining in state S_m^Ω , which is given by

$$C_m^\Omega(m | X_0^t) = c \sum_{n \in \Omega} P(T_n \leq t | X_0^t) + E[J_{t+1}(S_m^\Omega | X_0^{t+1}) | X_0^t], \quad (8)$$

where the incurred cost for time slot t is the sum of miss detection probabilities of all active channels.

$C_m^\Omega(n | X_0^t)$ is the cost to go for transiting to state S_n^Ω without stopping the communication over channel m , which is given by

$$C_m^\Omega(n | X_0^t) = c \sum_{s=t}^{t+d_s} \sum_{n \in \Omega} P(T_n \leq s | X_0^t) + E[J_{t+1+d_s}(S_n^\Omega | X_0^{t+d_s+1}) | X_0^t], \quad (9)$$

where the the incurred cost for time slot t is the sum of miss detection probabilities of all active channels during the blind period (recall that the spectrum sensor cannot sense any channel during this blind period).

$\tilde{C}_m^\Omega(n | X_0^t)$ is the cost of jumping to state $S_n^{\Omega'}$ after stopping the communication on channel m , which is given by

$$\begin{aligned} \tilde{C}_m^\Omega(n | X_0^t) &= c \sum_{s=t}^{t+d_s} \sum_{n \in \Omega'} P(T_n \leq t | X_0^t) \\ &+ E[J_{t+1+d_s}(S_n^{\Omega'} | X_0^{t+d_s+1}) | X_0^t] \\ &+ P(T_m > t | X_0^t), \end{aligned} \quad (10)$$

where $\Omega' = \Omega / \{m\}$ and incurred cost at time slot t is the sum of the false alarm probability for channel m and miss detection probabilities for other active channels.

The cost-to-go functions can be computed in a backward manner, that is, begin from J_Γ and compute J_t based on obtained J_{t+1} , until J_1 .

4.2. Sufficient Statistics. In this subsection, we find sufficient statistics for the cost-to-go functions.

4.2.1. Sufficiency. Notice that, in (6)–(10), the cost-to-go functions are dependent on observations X_0^t , which consume prohibitive amount of memory. Using a similar proof to that of in [21, Proposition 3] (for completeness, we provide the proof in Appendix B), we obtain the following proposition, which states that we need only keep *a posteriori* probabilities in the memory. (Since we have only partial information about the state of primary users, it is essentially a partially observable Markov decision process (POMDP). In many circumstances of POMDP, we can use the belief of the state (the *a posteriori* probabilities in our context) as the system state, thus converting the POMDP problem to a completely observable problem.)

Proposition 1. *The a posteriori probabilities $\{P(T_m \leq t | X_0^t)\}_{m=1,\dots,M}$ are sufficient statistics for the cost-to-go functions in (6)–(10).*

Therefore, we can update the *a posteriori* probabilities $\{P(T_m \leq t | X_0^t)\}_{m=1,\dots,M}$ for each new observation, instead of keeping all observations in memory. This requires only constant amount of memory.

4.2.2. Computation of A Posteriori Probabilities. The following proposition provides a formula to compute the *a posteriori* probability $P(T_n \leq t | X_0^t)$. The proof is given in Appendix C.

Proposition 2. *The a posteriori probability $P(T_n \leq t | X_0^t)$ for frequency channel n is given by*

$$\begin{aligned} P(T_n \leq t | X_0^t) &= \frac{\sum_{s=0}^t \left(\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s) \right)}{\sum_{s=0}^\infty \left(\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s) \right)}. \end{aligned} \quad (11)$$

For evaluating the *a posteriori* probability $P(T_n \leq t | X_0^t)$ recursively, we define the following quantity:

$$\begin{aligned} a_t^n(X_0^t) &\triangleq \prod_{r=0, r \in O_n}^t p_{0n}(X_r) \\ &= \begin{cases} a_{t-1}^n(X_0^{t-1}) p_{0n}(X_t), & \text{if } t \in O_n, \\ a_{t-1}^n(X_0^{t-1}), & \text{if } t \notin O_n. \end{cases} \end{aligned} \quad (12)$$

Based on the definition of $a_t^n(X_0^t)$ in (12), the numerator and denominator of the *a posteriori* probability $P(T_n \leq t | X_0^t)$ in (11) are given by

$$\begin{aligned} b_t^n(X_0^t) &\triangleq \text{numerator of (11)} \\ &= \begin{cases} b_{t-1}^n(X_0^{t-1}) p_{1n}(X_t) \\ + a_{t-1}^n(X_0^{t-1}) p_{1n}(X_t) p_e(t), & \text{if } t \in O_n, \\ b_{t-1}^n(X_0^{t-1}) + a_{t-1}^n(X_0^{t-1}) p_e(t), & \text{if } t \notin O_n. \end{cases} \end{aligned} \quad (13)$$

$c_t^n(X_0^t) \triangleq \text{denominator of (11)}$

$$\begin{aligned} &= \begin{cases} b_{t-1}^n(X_0^{t-1}) p_{1n}(X_t) + a_{t-1}^n(X_0^{t-1}) \\ \times \left(p_{1n}(X_t) p_e(t) + p_{0n}(X_t) \sum_{s=t+1}^\infty p_e(s) \right), & \text{if } t \in O_n, \\ b_{t-1}^n(X_0^{t-1}) + a_{t-1}^n(X_0^{t-1}) \times \left(\sum_{s=t}^\infty p_e(s) \right), & \text{if } t \notin O_n, \end{cases} \end{aligned} \quad (14)$$

where the numerator $b_t^n(X_0^t)$ is also computed recursively. The initialization of $a_t^n(X_0^t)$ and $b_t^n(X_0^t)$ is given by $a_t^n(X_0^{-1}) = 1$ and $b_t^n(X_0^{-1}) = 0$. The detailed derivation of (13) and (14) is given in Appendix D.

4.2.3. Prediction of Future Probabilities. Since the *a posteriori* probabilities $P(T_n \leq t \mid X_0^t)$ are sufficient statistics, we can rewrite the cost-to-go function $J_t(S_m^\Omega \mid X_0^t)$ as $J_t(S_m^\Omega \mid \mathbf{p}_t)$, where

$$(\mathbf{p}_t)_m = \begin{cases} P(T_m \leq t \mid X_0^t), & \text{if } m \in \Omega, \\ 0, & \text{if } m \notin \Omega, \end{cases} \quad (15)$$

in the remainder of this paper.

Conditioned on \mathbf{p}_t , the n th element of \mathbf{p}_{t+1} is given by

$$\begin{aligned} P(T_n \leq t+1 \mid X_0^t) &= P(T_n \leq t \mid X_0^t) + P(T_n = t+1 \mid X_0^t) \\ &= (\mathbf{p}_t)_n + \frac{P(X_0^t \mid T_n = t+1)P(T_n = t+1)}{P(X_0^t)} \\ &= (\mathbf{p}_t)_n + \frac{P(X_0^t \mid T_n > t)p_e(t+1)}{P(X_0^t)} \\ &= (\mathbf{p}_t)_n + \frac{P(X_0^t \mid T_n > t)P(T_n > t)p_e(t+1)}{P(X_0^t)P(T_n > t)} \\ &= (\mathbf{p}_t)_n + \frac{P(T_n > t \mid X_0^t)p_e(t+1)}{\sum_{s=t+1}^{\infty} p_e(s)} \\ &= (\mathbf{p}_t)_n + \rho(1 - (\mathbf{p}_t)_n). \end{aligned} \quad (16)$$

Using similar argument, we can show that for all $s > 0$,

$$P(T_n \leq t+s \mid X_0^t) = (\mathbf{p}_t)_n + (1 - (1 - \rho)^s)(1 - (\mathbf{p}_t)_n). \quad (17)$$

5. Infinite Horizon Case

Although we have obtained the cost-to-go functions and an efficient algorithm for computing the *a posteriori* probabilities, the assumption of limited observation period is unreasonable for practical systems; moreover, the cost-to-go functions are distinct for different time slots, thus requiring prohibitive amount of memory for storing the corresponding control policies when Γ is large. Therefore, in this section, we simplify the cost-to-go functions by considering infinite horizon case, that is, extending the limited time period to an infinite one. We first show that the cost-to-go functions converge to a function independent of time and then study their properties for further simplification.

5.1. Convergence. We first obtain the following proposition, which eliminates the dependency of cost-to-go functions on time. The proof is given in Appendix E.

Proposition 3. *As $\Gamma \rightarrow \infty$, one has*

$$J_t(S_m^\Omega \mid \mathbf{p}_t) \rightarrow J(S_m^\Omega \mid \mathbf{p}_t), \quad \forall t, \quad (18)$$

where $J(S_m^\Omega \mid \mathbf{p}_t)$ is the cost-to-go function in the infinite horizon case.

Therefore, one can focus on studying the infinite horizon cost-to-go function $J(S_m^\Omega \mid \mathbf{p}_t)$, thus reducing the number of cost-to-go functions from $\Gamma \times$ the number of states to the number of states.

5.2. Properties. For further exploiting the structure of DP, we study the properties of $J(S_m^\Omega \mid \mathbf{p}_t)$.

Symmetry. Since frequency channels are assumed to be symmetric (if different channels have different probabilities of primary user emergence, the symmetry is broken and we cannot simplify the cost-to-go functions), we have

$$J(S_m^{\Omega_1} \mid \mathbf{p}_t) = J(S_n^{\Omega_2} \mid \mathbf{p}_t), \quad (19)$$

if $|\Omega_1| = |\Omega_2|$, $(\mathbf{p}_t)_m = (\mathbf{p}'_t)_n$ and \mathbf{p}'_t is a permutation of the elements in \mathbf{p}_t . Then, we can rewrite $J(S_m^{\Omega_1} \mid \mathbf{p}_t)$ as $J_m^k(\mathbf{p}_t)$, where m indicates the frequency channel being sensed, and $k = |\Omega|$ is the number of frequency channels being used. Moreover, without loss of generality, we can assume that channel 1 is being monitored and need to study only $J_1^k(\mathbf{p}_t)$ due to symmetry.

Then the cost-to-go function in (7) can be rewritten as

$$\begin{aligned} J_1^k(\mathbf{p}_t) &= \min \left(c \|\mathbf{p}_t\|_1 + \tilde{J}_n^k(\mathbf{p}_t), \right. \\ &\quad \left. c(A \|\mathbf{p}_t\|_1 + B \|\mathbf{p}_t\|_0) + \min_{n \neq 1} \tilde{J}_n^{k-1}(\mathbf{p}_t), \right. \\ &\quad \left. c(A \|\mathbf{p}_t^1\|_1 + B \|\mathbf{p}_t^1\|_0) \right. \\ &\quad \left. + \min_{n \neq 1} \tilde{J}_n^{k-1}(\mathbf{p}_t) + 1 - (\mathbf{p}_t)_m \right), \end{aligned} \quad (20)$$

where

$$\begin{aligned} \tilde{J}_m^k(\mathbf{p}_t) &= E \left[J_m^k(\mathbf{p}_{t+1}) \mid \mathbf{p}_t \right], \\ A &= \frac{1 - (1 - \rho_e)^{d_s+1}}{\rho_e}, \\ B &= d_s - \frac{1 - (1 - \rho_e)^{d_s+1}}{\rho_e}. \end{aligned} \quad (21)$$

Note that $A \|\mathbf{p}_t\|_1 + B \|\mathbf{p}_t\|_0$ corresponds to $\sum_{s=t}^{t+d_s} \sum_{n \in \Omega} P(T_n \leq s \mid X_0^t)$ in (9) and $A \|\mathbf{p}_t^1\|_1 + B \|\mathbf{p}_t^1\|_0$ corresponds to $\sum_{s=t}^{t+d_s} \sum_{n \in \Omega, n \neq 1} P(T_n \leq s \mid X_0^t)$ in (10). \mathbf{p}_t^m is obtained by setting the m th element in \mathbf{p}_t to 0.

Argmin. If transiting to another frequency channel, the secondary user should always choose the frequency channel having the largest *a posteriori* probability, that is,

$$\arg \min_{n \neq 1} J_n^k(\mathbf{p}_t) = \arg \max_{n \neq 1} (\mathbf{p}_t)_n. \quad (22)$$

Therefore, the computation of cost-to-go functions can be simplified to

$$\begin{aligned}
J_1^k(\mathbf{p}_t) &= \min\left(c\|\mathbf{p}_t\|_1 + \tilde{J}_1^k(\mathbf{p}_t), \right. \\
&\quad c(A\|\mathbf{p}_t\|_1 + B\|\mathbf{p}_t\|_0) + \tilde{J}_1^k(\pi(\mathbf{p}_t)), \\
&\quad \left. c(A\|\mathbf{p}_t^1\|_1 + B\|\mathbf{p}_t^1\|_0) + \tilde{J}_1^{k-1}(\pi(\mathbf{p}_t)) \right. \\
&\quad \left. + 1 - (\mathbf{p}_t)_1\right), \tag{23}
\end{aligned}$$

where π is an operator that switches the elements belonging to frequency channel 1 and the frequency channel given by (22), that is,

$$\pi(\mathbf{x}) = \begin{cases} (\pi(\mathbf{x}))_1 = \max_{n \neq 1}(\mathbf{x})_n, \\ (\pi(\mathbf{x}))_n = (\mathbf{x})_1, & \text{if } n = \arg \max_{n \neq 1}(\mathbf{x})_n, \\ (\pi(\mathbf{x}))_n = (\mathbf{x})_n, & \text{if } n \neq \arg \max_{n \neq 1}(\mathbf{x})_n. \end{cases} \tag{24}$$

6. Heuristic Approximation

The probability \mathbf{p}_t is continuous, thus resulting in infinite numbers of cost-to-go functions $J_1^k(\mathbf{p}_t)$. Therefore, we need to discretize the probability \mathbf{p}_t into f intervals for numerical computation. It is easy to verify that the number of cost-to-go functions is given by $\sum_{m=1}^M f^m = (f^{M+1} - f)/(f - 1)$ (when there are still m active channels, there are f^m possibilities for $J_1^k(\mathbf{p}_t)$). When the number of frequency channels is large, we face the curse of dimensions for numerically computing the cost-to-go functions in (23). For example, when $f = 10$ and $M = 10$, we need to consider around 10^{10} cost-to-go functions. Therefore, we need approximations to simplify DP. There have been plenty of studies on approximate DP [22–24]. In this paper, we combine the philosophies of Limited Lookahead Policy (LLP), which truncates the time horizon by looking ahead only a small number of stages, and Certainty Equivalent Control (CEC), which replaces random variables with their expectations, in [19].

- (i) LLP: intuitively, in the near future, the first two most possibly changed frequency channels are the one being monitored and the one not being monitored but having the largest *a posteriori* probability (if there is a tie, we can choose one randomly). For simplicity, we assume that they are channels 1 and 2, respectively. Applying the philosophy of LLP, we consider only these two frequency channels and do not consider any other frequency channels.
- (ii) CEC: using the philosophy of CEC, we convert the stochastic control problem into a deterministic one, that is, considering the expectations of change times, $\bar{T}_n^t \triangleq E[T_n | X_0^t]$, to be the true values.

The following proposition provides expressions for the expected changing time.

Proposition 4. For any n , the expected changing time of channel n is given by

$$\begin{aligned}
\bar{T}_n^t &= \frac{\sum_{s=0}^t \left(s \prod_{r=0, r \in O_1}^{s-1} P_{0n}(X_r) \prod_{r=s, r \in O_1} P_{1n}(X_r) p_e(s) \right)}{\sum_{s=0}^{\infty} \left(\prod_{r=0, r \in O_n}^{s-1} P_{0n}(X_r) \prod_{r=s, r \in O_n}^t P_{1n}(X_r) p_e(s) \right)} \\
&\quad + \left(t + \frac{1}{\rho} \right) (1 - (\mathbf{p}_t)_n). \tag{25}
\end{aligned}$$

Obviously, the denominator of the first term in (25) can be computed using (14). The corresponding numerator can be computed recursively (similar to (13)) as follows:

$$\begin{aligned}
d_t^n(X_0^t) &\triangleq \text{numerator of (25)} \\
&= \begin{cases} d_{t-1}^n(X_0^{t-1}) P_{1n}(X_t) \\ + t d_{t-1}^n(X_0^{t-1}) P_{1n}(X_t) p_e(t), & \text{if } t \in O_n, \\ d_{t-1}^n(X_0^{t-1}) + t d_{t-1}^n(X_0^{t-1}) p_e(t), & \text{if } t \notin O_n. \end{cases} \tag{26}
\end{aligned}$$

For compensating the false alarm probability and state transition time d_s , we do the following adjustments for channels 1 and 2:

$$\tilde{T}_1^t = \bar{T}_1^t + \frac{1}{c} (1 - (\mathbf{p}_t)_1) \tag{27}$$

$$\tilde{T}_2^t = \bar{T}_2^t + \frac{1}{c} (1 - (\mathbf{p}_t)_2) + d_s. \tag{28}$$

Note that $1/c$ is used to convert the penalty of false alarm to detection delay, and d_s is applied to channel 2 since there is no blind period if we continue to monitor channel 1.

Then, a heuristic decision of state transition is given by (as illustrated in Figure 4) as the following.

- (i) Case 1: if $\tilde{T}_1^t \leq t$, stop using frequency channel 1 and switch to monitor frequency channel 2.
- (ii) Case 2: if $\tilde{T}_1^t > t$ and $\tilde{T}_1^t > \tilde{T}_2^t$, continue using frequency channel 1 and switch to monitor frequency channel 2.
- (iii) Case 3: if $\tilde{T}_1^t > t$ and $\tilde{T}_1^t \leq \tilde{T}_2^t$, keep monitoring frequency channel 1.

7. Numerical Results

In this section, we use numerical simulation results to evaluate the performance of the proposed selective quickest spectrum sensing. The following configurations are used for all simulations.

- (i) We assume $M = 2$, that is, there are two frequency channels used by the secondary user.
- (ii) We consider sensed power (in dB scale) as observation which satisfies Gaussian distribution, that is,

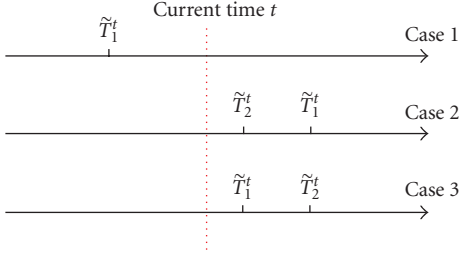


FIGURE 4: Illustration of three cases in the heuristic strategy.

$H_0 : X_t \sim \mathcal{N}(P_0, \sigma_n^2)$ and $H_1 : X_t \sim \mathcal{N}(P_1, \sigma_n^2)$, where P_0 and P_1 are the expected receive power (in dB) with and without primary users, respectively, and σ_n^2 is the variance of measurement error incurred by fading, noise and interference. We assume that the signal-to-noise ratio (SNR) is 10 dB. Note that the normality assumption is mainly for simplicity of simulation and is correct if log-normal distributed shadow fading is considered. Such a normality assumption has been used in many other publications, for example, [25]. It is also straightforward to incorporate other possible observation distributions, for example, incorporating Rayleigh or Ricean fading and thermal noise, into the framework of selective quickest spectrum sensing.

(iii) d_s is set to 10 time slots.

Each simulation statistic is obtained from 1000 realizations of the spectrum sensing procedure.

7.1. Discretized DP. For computing the cost-to-go functions, we discretize the *a posteriori* probabilities by dividing the range (between 0 and 1) of each probability into 30 equal length intervals. 100 iterations are used to compute these cost-to-go functions. Then, the obtained control policy is applied to the spectrum sensing. Note that the computation of control policy is offline and does not affect the realtime operation of the secondary user.

Figure 5 shows the trace of control action in one realization of the spectrum sensing process. The upper slashed black curve represents the current frequency channel being monitored. Four events are labeled in the figure:

- (i) event 1: primary user emerges in channel 2;
- (ii) event 2: primary user emerges in channel 1;
- (iii) event 3: the secondary user quits channel 2;
- (iv) event 4: the secondary user quits channel 1.

The *a posteriori* probabilities $P(T_i \leq t | X_0^t)$, $i = 1, 2$, are both plotted in the figure. In the figure, the procedure of spectrum sensing is as follows:

- (1) at the very beginning, both *a posteriori* probabilities are small and the secondary user switches to channel 2 from channel 1;
- (2) during the blind period, the node cannot monitor any frequency channel; then the secondary user begins to monitor channel 2;

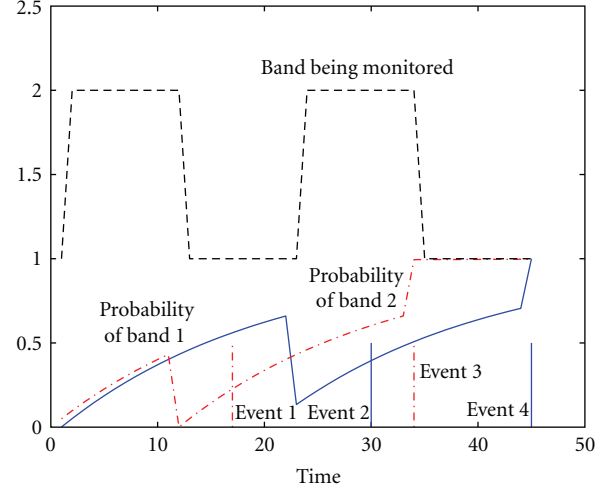


FIGURE 5: An example of control action trace.

- (3) when the *a posteriori* probability of channel 1 (black solid curve) becomes much larger than that of channel 2, the secondary user switches back to channel 1;
- (4) when the *a posteriori* probability of channel 2 becomes much larger than that of channel 1, the secondary user switches back to channel 2; after the blind period, the node detects the change of channel 2;
- (5) the secondary user quits channel 2 and begins to monitor channel 1; after the blind period, it detects the change of channel 1.

Figure 6 shows the cumulative distribution function (CDF) of detection delay when $\rho = 0.05, 0.1, 0.15$, where we set $c = 0.05$. We observe that the performance is improved when ρ is increased. An intuitive explanation is that the emergence of the primary users is less random when ρ is larger.

Figure 7 shows the tradeoff between false alarm rate and detection delay ARL (recall that the detection delay ARL is defined as $E[(t_m - T_m)^+]$), where we set $\rho = 0.05, 0.15$. We change the weighting factor c to generate curves characterizing different tradeoffs between false alarm and miss detection and observe that the tradeoff curve is much better when $\rho = 0.15$.

7.2. Approximate DP. Figures 8 and 9 show the performance (CDF of detection delay and tradeoff curves) of approximate DP in Section 6. In Figure 9, the approximate DP even outperforms the discretized DP at some points; for example, for $\rho = 0.05$ and detection delay ARL equaling 8, the false alarm rate of approximate DP is smaller than that of the discretized DP. Note that this does not contradict the optimality of DP since the DP uses discretized probabilities while the approximate DP does not.

Although the approximate DP achieves good performance when false alarm rate is small, our simulation shows that it cannot achieve low detection delay ARL even if we set

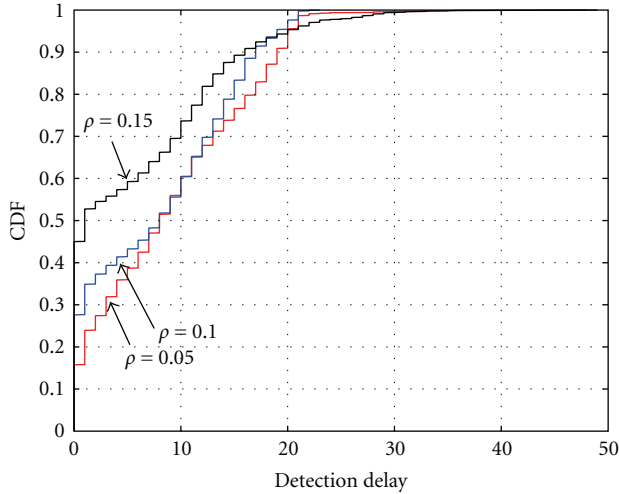


FIGURE 6: CDF of detection delay for $\rho = 0.05, 0.1, 0.15$ when discretized DP is used.

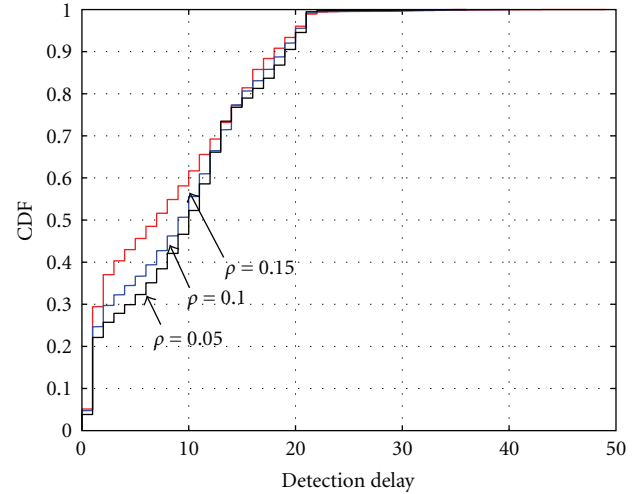


FIGURE 8: CDF of detection delay for $\rho = 0.05, 0.1, 0.15$ when approximate DP is used.

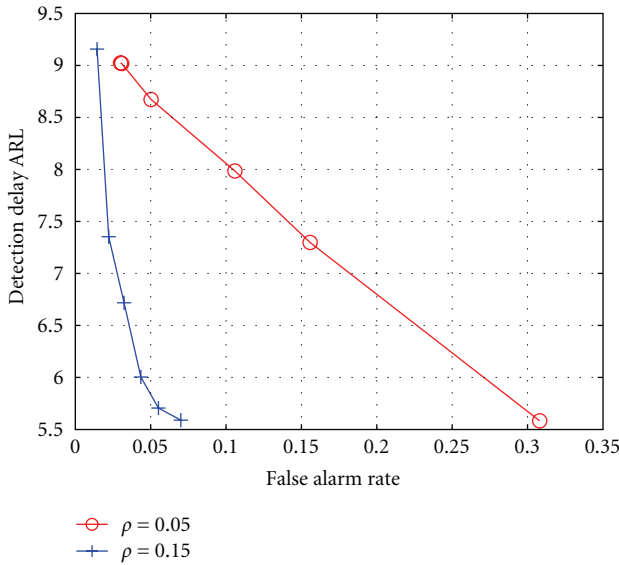


FIGURE 7: Tradeoff between false alarm rate and detection delay ARL when discretized DP is used.

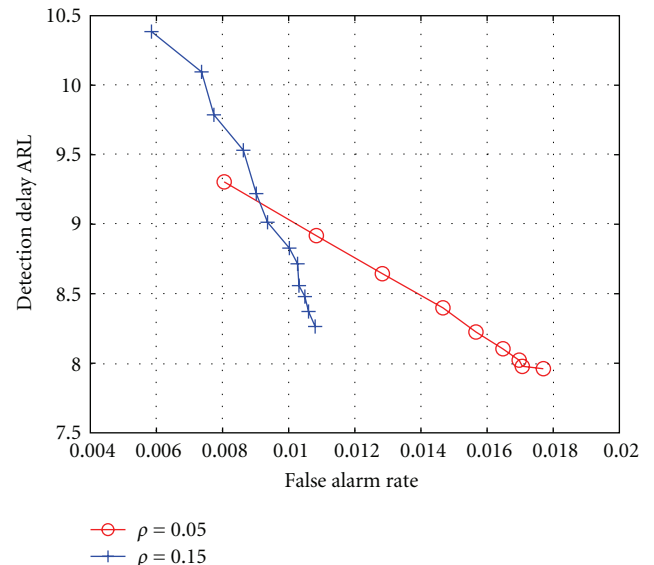


FIGURE 9: Tradeoff between false alarm rate and detection delay ARL when approximate DP is used.

the weighting factor c to a large number (i.e., emphasizing more on the penalty of detection delay). For the optimal DP, the controller tends to close the current frequency channel immediately to avoid the penalty of detection delay if c diverges to infinity. However, when we set $c = \infty$ in the approximate DP, the only effect is that the second terms in both (27) and (28) vanish, which does not necessarily imply stopping transmitting over the current frequency channel immediately. Therefore, the proposed approximate DP is less flexible than the optimal (or discretized) one.

8. Conclusions and Open Problems

We have applied the framework of DP to the problem of selective quickest spectrum sensing with blind period in

multichannel cognitive radio systems. A cost-to-go function based control policy is established for the restless watchdog to achieve tradeoff between detection delay and false alarm. *A posteriori* probabilities of primary user emergence are used as sufficient statistics with efficient recursive computation formulas. We have proposed a heuristic and approximate algorithm to avoid the curse of dimensions in DP. Numerical simulation shows that both the DP and approximate DP frameworks yield good performance for spectrum sensing.

There are still many open problems for the selective spectrum sensing. Major open problems include (a) when the statistics of primary user's activity are unknown or change in time, how to learn the optimal strategy adaptively? (b) when multiple secondary users exist, how to handle the competition among them?

Appendices

A. Dynamic Programming

In this section, we briefly introduce the principle of DP, making this paper self-contained. Consider a discrete-time Markovian system, whose evolution is described by

$$s_{t+1} = f(s_t, u_t, w_t), \quad (\text{A.1})$$

where f is a deterministic function, s_t is the state at time t , u_t is a legal action when the state is s_t , and w_t is some random perturbation. Consider a finite time interval $[1, T]$. The cost function of the system is given by

$$J = \sum_{t=1}^T E[c(s_t, u_t, w_t)], \quad (\text{A.2})$$

where c is a function mapping to a real number.

Following the basic idea of DP, that is, decomposing a problem into subproblems, we define cost-to-go function (it is also called value function if we consider reward instead of cost), $J_t(s)$, that is, the expected cost after time $t-1$ provided that $s_t = s$, which is given by

$$J_t(s) = \sum_{\tau=t}^T E[c(s_\tau, u_\tau, w_\tau) \mid s_t = s]. \quad (\text{A.3})$$

Denoting by the optimal (equivalently, minimal) cost-to-go function by $J_t^*(s_t)$, we have Bellman's Equation, which is given by

$$J_t^*(s) = \min_{u_t} E[c(s, u_t, w_t) + J_{t+1}^*(s_{t+1})], \quad (\text{A.4})$$

and the corresponding optimal control policy can be obtained by

$$\mu_t^*(s) = \arg \min_{u_t} E[c(s, u_t, w_t) + J_{t+1}^*(s_{t+1})]. \quad (\text{A.5})$$

B. Proof of Proposition 1

Proof. We do induction on time slot t . Due to (6), $\{P(T_n \leq \Gamma \mid X_0^\Gamma)\}_{n \in \Omega}$ is sufficient for $J_\Gamma(S_m^\Omega \mid X_0^\Gamma)$. Then, suppose that the *a posteriori* probabilities $\{P(T_n \leq t+1 \mid X_0^{t+1})\}_{n \in \Omega}$ are sufficient for the cost-to-go function $J_{t+1}(S_m^\Omega \mid X_0^{t+1})$. Now, we consider time slot t . Due to (17), $P(T_n \leq t+s \mid X_0^t)$ is a function of $P(T_n \leq t \mid X_0^t)$, for all $s \geq 0$. Then, (7) implies that $J_t(S_m^\Omega \mid X_0^t)$ depends on only $P(T_n \leq t \mid X_0^t)$ according to the induction assumption. This concludes the proof. \square

C. Proof of Proposition 2

Proof. It is easy to verify that the probability conditioned on known times of primary users' emergence on all channels $P(X_0^t \mid T_1 = s_1, \dots, T_M = s_M)$ is given by (recall that O_m is the set of time slots in which channel m is sensed)

$$P(X_0^t \mid T_1 = s_1, \dots, T_M = s_M) = \prod_{m=1}^M \prod_{r=0, r \in O_m}^{s_m-1} p_{0m}(X_r) \prod_{r=s_m, r \in O_m}^t p_{1m}(X_r) p_e(s_m). \quad (\text{C.1})$$

Similarly, we have

$$\begin{aligned} P(X_0^t \mid T_1 = s_1) &= \prod_{r=0, r \in O_1}^{t_1-1} p_{01}(X_r) \prod_{r=t_1, r \in O_1}^t p_{11}(X_r) \\ &\times \sum_{s_2=0}^{\infty} \cdots \sum_{s_m=0}^{\infty} \prod_{m=2}^M \prod_{r=0, r \in O_m}^{s_m-1} p_{0m}(X_r) \\ &\times \prod_{r=s_m, r \in O_m}^t p_{1m}(X_r) p_e(s_m). \end{aligned} \quad (\text{C.2})$$

Based on the above results, the unconditional probability $P(X_0^t)$ is given by

$$\begin{aligned} P(X_0^t) &= \sum_{s_1=0}^{\infty} \cdots \sum_{s_m=0}^{\infty} P(X_0^t \mid T_1 = s_1, \dots, T_M = s_M) \\ &\times P(T_1 = s_1, \dots, T_M = s_M) \\ &= \sum_{s_1=0}^{\infty} \cdots \sum_{s_m=0}^{\infty} \prod_{m=1}^M \prod_{r=0, r \in O_m}^{s_m-1} p_{0m}(X_r) \\ &\times \prod_{r=s_m, r \in O_m}^t p_{1m}(X_r) p_e(s_m), \end{aligned} \quad (\text{C.3})$$

where s_m stands for the possible time when primary users emerge on channel m .

On applying Bayes formula, the *a posteriori* probability $P(T_n \leq t \mid X_0^t)$ for frequency channel n is given by

$$\begin{aligned} P(T_n \leq t \mid X_0^t) &= \frac{P(X_0^t, T_n \leq t)}{P(X_0^t)} \\ &= \frac{P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n \leq t)}{P(\{X_\tau\}_{\tau \in O_n, \tau \leq t})} \\ &= \frac{\sum_{s=0}^t \left(\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s) \right)}{\sum_{s=0}^{\infty} \left(\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s) \right)}. \end{aligned} \quad (\text{C.4})$$

This concludes the proof. \square

D. Proof of Equations (13) and (14)

Proof. We first show (13). From the proof of Proposition 2, we know

$$b_t^*(X_0^t) = P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n \leq t). \quad (\text{D.1})$$

When $t \in O_n$, we have

$$\begin{aligned}
& P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n \leq t) \\
&= P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n \leq t-1) \\
&\quad + P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n = t) \\
&= P(\{X_\tau\}_{\tau \in O_n, \tau \leq t-1}, T_n \leq t-1)P(X_t | T_n \leq t-1) \\
&\quad + P(\{X_\tau\}_{\tau \in O_n, \tau \leq t-1} | T_n = t)P(X_t | T_n = t)P(T_n = t) \\
&= b_{t-1}^n(X_0^{t-1})p_{1n}(X_t) \\
&\quad + a_{t-1}^n(X_0^{t-1})p_{1n}(X_t)p_e(t),
\end{aligned} \tag{D.2}$$

where we applied $P(X_t | T_n \leq t-1) = P(X_t | T_n = t) = p_{1n}(X_t)$ and the definition of $a_{t-1}^n(X_0^{t-1})$.

When $t \notin O_n$, the derivation is the same except that the $p_{1n}(X_t)$ is not taken into account. This concludes the proof of (13).

Now we show (14). From the proof of Proposition 2, we know

$$c_t^n(X_0^t) = P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}). \tag{D.3}$$

Then, when $t \in O_n$, we have

$$\begin{aligned}
& P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}) \\
&= P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n \leq t-1) \\
&\quad + P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n = t) \\
&\quad + P(\{X_\tau\}_{\tau \in O_n, \tau \leq t}, T_n > t),
\end{aligned} \tag{D.4}$$

where the first two terms are the same as those of $b_t^n(X_0^t)$. The last term is equal to $a_{t-1}^n(X_0^{t-1})p_{0n}(X_t) \sum_{s=t+1}^{\infty} p_e(s)$ since the observation distribution is p_{0n} before time $t+1$ when $T_n > t$.

We can also obtain $c_t^n(X_0^t)$ by ignoring $p_{0n}(X_t)$ and $p_{1n}(X_t)$ when $t \notin O_n$. This concludes the proof. \square

E. Proof of Proposition 3

Proof. Let $\Gamma_1 \geq \Gamma_2 \geq 0$, we have

$$J_t^{\Gamma_1}(S_m^\Omega | \mathbf{p}_t) > J_t^{\Gamma_2}(S_m^\Omega | \mathbf{p}_t), \quad \forall t, \tag{E.1}$$

where the superscripts Γ_1 and Γ_2 indicate the final time, since larger time interval means more possible false alarms and detection delays.

However, $J_t^{\Gamma}(S_m^\Omega | \mathbf{p}_t)$ is upper bounded by M since it is smaller than or equal to the cost of the simple strategy that the secondary user claims the emergence of primary users over all channels at time 0. Due to (3), the cost of the strategy is M since $P(T_m > 0) = 1$ and $P(T_m \leq 0) = 0$.

Therefore, $J_t^{\Gamma}(S_m^\Omega | \mathbf{p}_t)$ is an upper bounded increasing function in Γ and thus converges as $\Gamma \rightarrow \infty$. This concludes the proof. \square

F. Proof of Proposition 4

Proof. Similar to the proof of Proposition 2 (in Appendix C), we obtain

$$\begin{aligned}
& P(T_n = s | X_0^t) \\
&= \frac{(\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s))}{\sum_{s=0}^{\infty} (\prod_{r=0, r \in O_n}^{s-1} p_{0n}(X_r) \prod_{r=s, r \in O_n}^t p_{1n}(X_r) p_e(s))},
\end{aligned} \tag{F.1}$$

when $s \leq t$.

When $T_n > t$ (this happens with probability $1 - (\mathbf{p})_n$), the conditional expectation is given by

$$E[T_n - t | X_0^t, T_n > t] = \frac{1}{\rho}, \tag{F.2}$$

where $1/\rho$ is the unconditional expectation of the time of primary users' emergence.

Then, we have

$$\begin{aligned}
\bar{T}_n^t &= \sum_{s=0}^t sP(T_n = s | X_0^t) \\
&\quad + (E[T_n - t | X_0^t, T_n > t] + t)P(T_n > t | X_0^t),
\end{aligned} \tag{F.3}$$

which concludes the proof. \square

Acknowledgments

This paper is supported by the National Science Foundation by grant CCF-0830451. Part of this paper has been presented in IEEE International Conference on Communications (ICC) in 2009.

References

- [1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proceedings of IEEE International Workshop Mobile Multimedia Communications*, pp. 3–10, 1999.
- [2] J. Mitola, *Cognitive Radio*, Licentiate Proposal, KTH, Stockholm, Sweden, 1998.
- [3] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," November 2002, <http://www.fcc.gov/sptf/reports.html>.
- [4] A. Sahai, R. Tandra, M. Mishra, and N. Hoven, "Fundamental design tradeoffs in cognitive radio systems," in *Proceedings of the 1st International Workshop on Technology and Policy for Accessing Spectrum (TAPAS '06)*, Boston, Mass, USA, August 2006.
- [5] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, pp. 79–89, 2007.
- [6] IEEE, *IEEE 802 LAM/MAN Standards Committee 802.22 WG on WRANS (Wireless Regional Area Networks)*, 2009.
- [7] C. R. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional requirements for the 802.22 wran standard," IEEE 802.22-05/0007r47, January 2006.
- [8] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG dynamic spectrum sharing field test results," in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pp. 676–684, 2007.

- [9] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 297–307, 2009.
- [10] L. Lai, H. E. Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: exploitation, exploration and competition," submitted to *IEEE/ACM Transactions on Networking*.
- [11] H. Robbins, "Some aspects of the sequential design of experiments," *American Mathematical Society*, vol. 58, pp. 527–535, 1952.
- [12] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of Whittle's index for dynamic multichannel access," submitted to *IEEE Transactions on Information Theory*.
- [13] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, 2007.
- [14] S. H. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," submitted to *IEEE Transactions on Information Theory*.
- [15] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS '08)*, pp. 203–208, Princeton, NJ, USA, 2008.
- [16] H. V. Poor and O. Hadjiladis, *Quickest Detection*, Cambridge University Press, Cambridge, UK, 2008.
- [17] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multichannel opportunistic access: structure, optimality and performance," to appear in *IEEE Transactions on Wireless Communications*.
- [18] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, USA, 1957.
- [19] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
- [20] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Transactions on Information Theory*, vol. 47, pp. 1657–1665, 2001.
- [21] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Transactions on Information Theory*, vol. 39, pp. 433–442, 1993.
- [22] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Boston, Mass, USA, 1996.
- [23] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Cambridge University Press, Cambridge, UK, 2007.
- [24] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, *Handbook of Learning and Approximate Dynamic Programming*, Wiley-IEEE Press, New York, NY, USA, 2004.
- [25] J. Unnikrishnan and V. V. Veeravalli, "Cooperative spectrum sensing and detection for cognitive radio," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 2972–2976, 2007.